GenePattern:

A platform for integrative genomics

Michael Reich

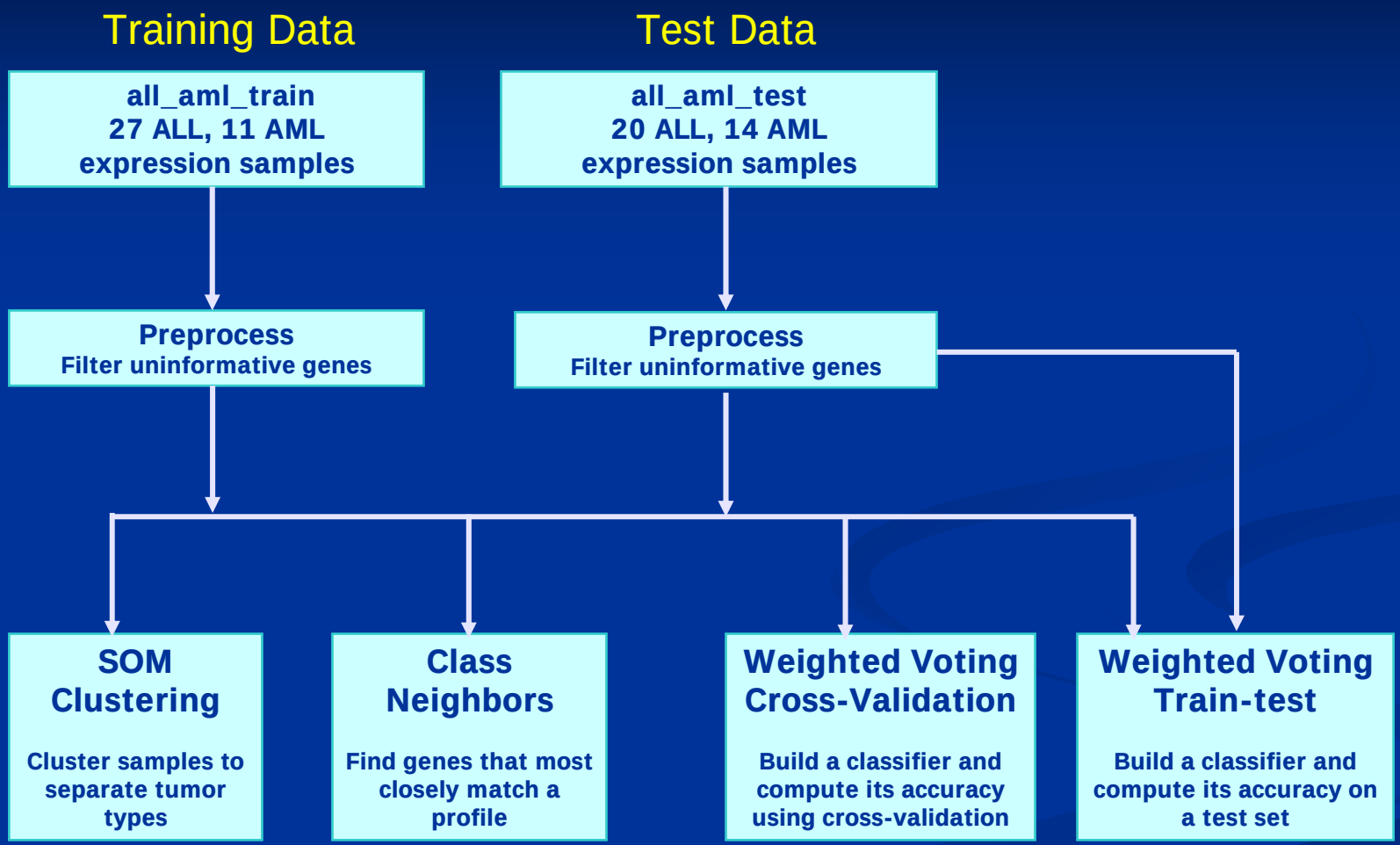
Broad Institute of MIT and Harvard

July 19, 2006

Challenges in genomic research

- The number of available research tools is growing exponentially.
- Published results of *in silico* research are not reproducible without extensive communication with the authors.
- Research teams comprise members from many disciplines and levels of computational sophistication.
- The research environment is dynamic and heterogeneous, with new tools being developed quickly in many different forms.

GenePattern initial DBP: ALL/AML Classification



Golub and Slonim et al., *Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression*, Science 1999

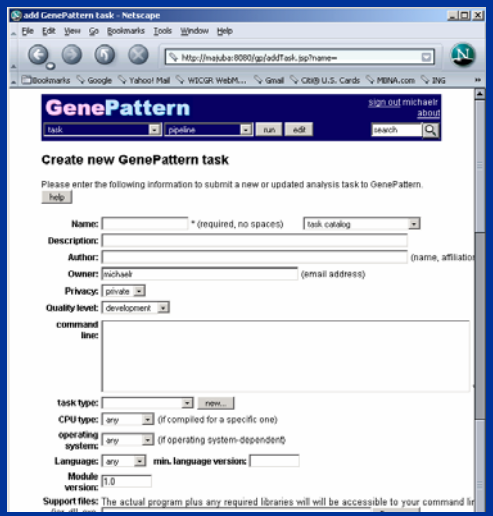
Design Objectives

- A comprehensive repository of analysis modules
- User interface accessible to users at all levels of computational sophistication
- Ability to chain tasks into pipelines for reproducible *in silico* research
- Ability to add new tools without programming
- Computing on a local machine or distributed among more powerful compute servers

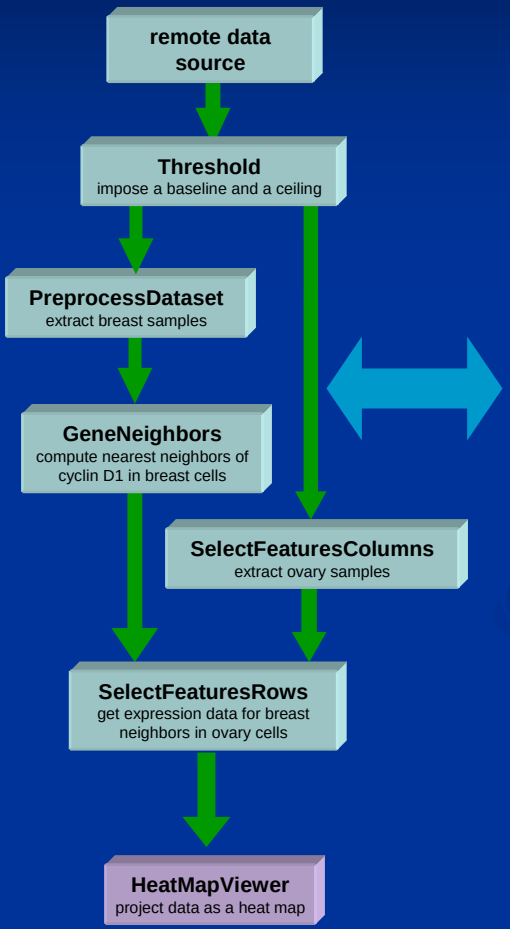
GenePattern: A platform for integrative genomics

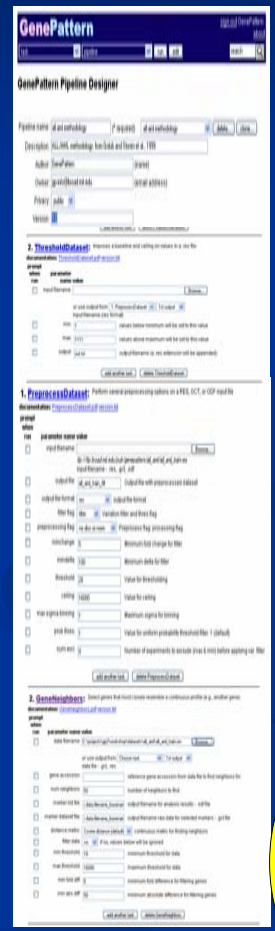
Module Repository

Task Integrator

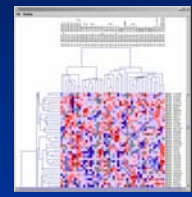


Pipeline Environment

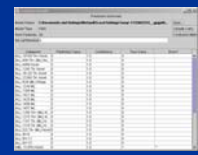




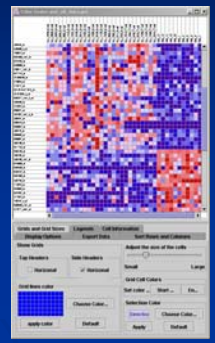
Graphical Environment



Bicluster

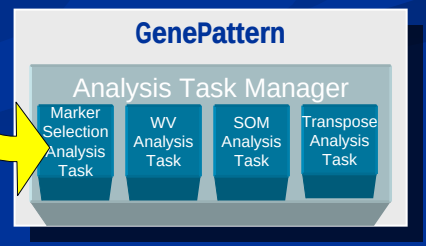


Prediction Results



Heat Map

Programming Environment



GenePattern Features

Comprehensive Module Repository

- ~70 modules: analysis, visualization, pipelines
- Expression, proteomic, sequence, *variation (SNP),* and *whole genome association* data
- Task suites

Analytic Reproducibility

- Easy, rapid sharing of methodologies via pipelines
- Versioning using Life Sciences Identifier (LSID)
- Executable history of all sessions
- Automatic pipeline generation from result files
- Executable research documents

Automatic Task Integration

- Add new modules without writing code
- Supports any command line callable code (language independent)

Multiple user interfaces

- Java client
- Web client
- Programmatic interfaces to Java, MATLAB, R, *Perl*

Local and Distributed Computing

- Laptop
- Client/Server
- Compute farm
- *Web Portal*

Interoperability

- caBIG
  - Interface to caArray
  - *caGrid service (Beta)*
- *geWorkbench (Beta)*

GenePattern Software

Release Information

- Initially released in March, 2004
- Current version 2.0.2, released July 2006
- Currently ~3000 users, 500+ organizations, ~90 countries

Availability

- Freely available
- Windows, Mac OS, and Linux platforms

Resources

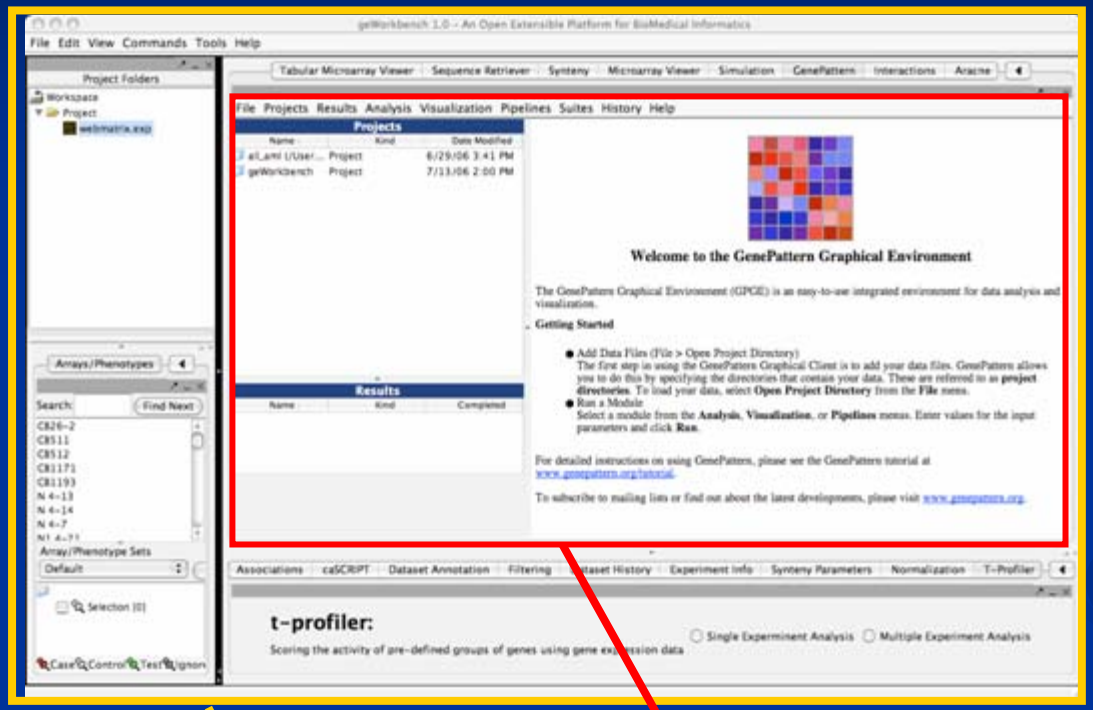
- http://www.broad.mit.edu/genepattern
- User workshops, documentation, email help desk, online user forum
- Reich et al. (2006) *Nature Genetics*

GenePattern is a winner of the
2005 BioIT World Best Practices Award

GenePattern/NCIBI Plan

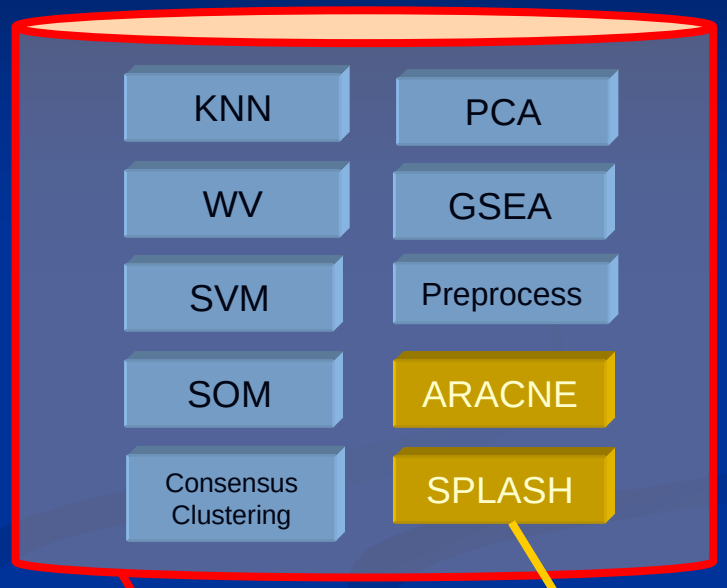
- Integration with NCIBI software to support driving biological problems and bioinformatics tools
- Delivery of live and Web-based GenePattern training workshops

GenePattern/geWorkbench Interoperability
with MAGNet NCBC and NCIBI NCBC



geWorkbench application

GenePattern UI plug-in



GenePattern module repository

geWorkbench modules

Execute GenePattern modules from within geWorkbench

Wrap geWorkbench modules as GenePattern tasks

Acknowledgements

Core Development

Josh Gould
Marc-Danie Nazaire
Jim Lerner
Ted Liefeld
Jim Robinson
David Twomey

Module Contributors

Gad Getz
Justin Lamb
D.R. Mani
Stefano Monti
Ken Ross
Aravind Subramanian

Todd R. Golub
Pablo Tamayo
Jill P. Mesirov, PI

www.broad.mit.edu/genepattern